

BAB III METODE PENELITIAN

3.1. Data Penelitian

Data penelitian yang digunakan pada penelitian ini bersumber dari Perusahaan yang bergerak di bidang percetakan. Data ini terdiri dari 770 dan 9 variabel diantaranya variable SAP, vendor, nilai awal, prestasi, pelanggaran, total nilai, spp ditawarkan, dan spp menawarkan.

3.2. Lokasi Penelitian dan Waktu Penelitian

Lokasi penelitian dilakukan pada perusahaan yang bergerak di bidang Percetakan. Adapun waktu pelaksanaan dimulai dari bulan Oktober 2022, Perinci waktu penelitian terdapat pada Table 3.1.

Tabel 3. 1 Waktu Penelitian

No	Kegiatan	2022				2023							
		11				12				1			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Analisis Masalah												
2	Pengumpulan Data												
3	Implementasi												
4	Pengujian												
5	Evaluasi												

3.3. Peralatan Penelitian

Peralatan yang digunakan pada penelitian ini meliputi perangkat keras dan perangkat lunak dengan rincian sebagai berikut :

1. Perangkat Lunak

a. Windows 11

Berfungsi sebagai sistem pengoprasian komputer dan laptop

b. Microsoft Office Excel 2010

Berfungsi untuk mengolah data teks dan angka.

a. *Microsoft Office Word 2010*

Berfungsi untuk mengolah, menyimpan data

b. *Google collab*

Berfungsi sebagai *coding environment* yang dijalankan melalui browser tanpa harus instalasi.

c. *Python*

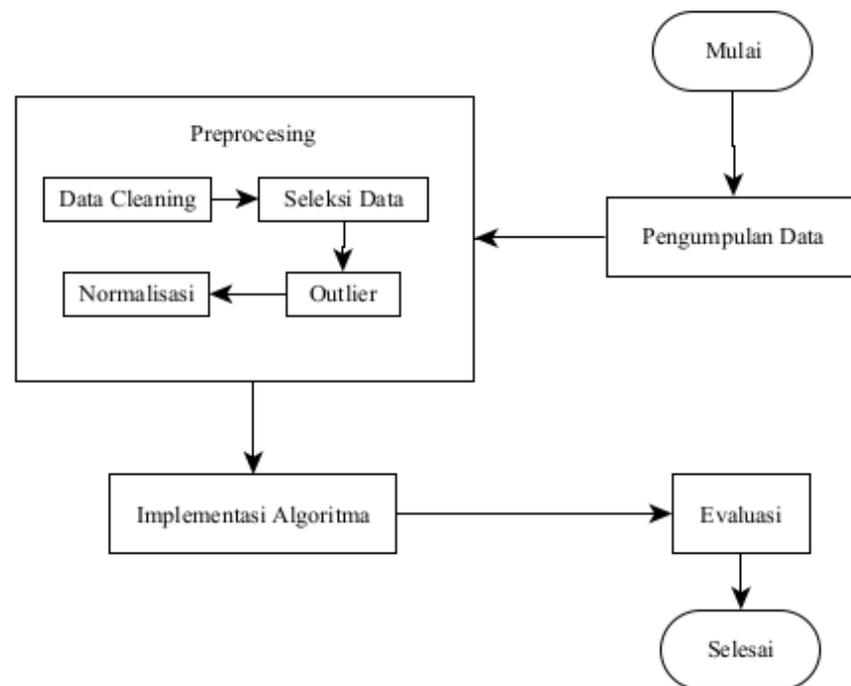
Sebagai bahasa yang digunakan pada penelitian ini

1. Perangkat Keras

Penelitian ini menggunakan perangkat keras dengan spesifikasi RAM 4 GB, *Processor Intel(R) Celeron(R) N4020 CPU @ 1.10GHz 1.10 GHz*, *64-bit operating system, x64-based processor*

1.1. Prosedur Penelitian

Tahapan prosedur penelitian dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Prosedur Penelitian

Berdasarkan Gambar 3.1 dapat dilihat bahwa penelitian ini terdiri dari enam tahap yang dilakukan secara berurutan, yaitu pengumpulan data, *preprocessing*, implementasi algoritma, pengolahan data dengan teknik *clustering* menggunakan perbandingan antara algoritma *K-Medoids* dan *Fuzzy c-means*, serta evaluasi dengan menguji validasi *cluster* yang dihasilkan oleh kedua algoritma tersebut.

1.1.1. Pengumpulan Data

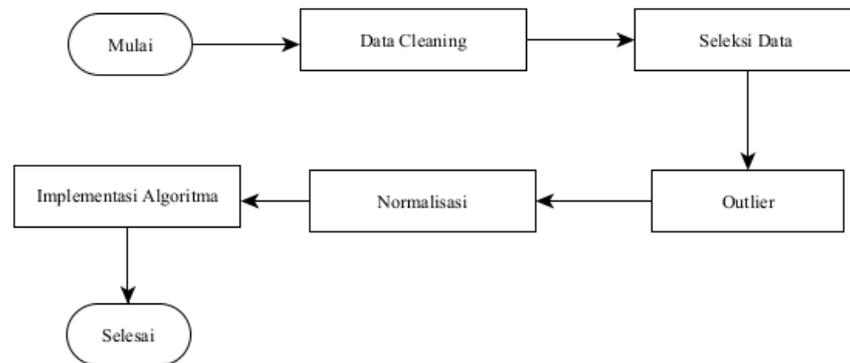
Setelah menentukan masalah yang akan diselesaikan dan membaca literatur terkait. Tahap selanjutnya mengumpulkan dan menyiapkan dataset, dalam proses ini maka dibutuhkan data yang sumbernya valid dan relevan. Maka untuk memperoleh data tersebut peneliti mengambil data dari perusahaan yang bergerak di bidang percetakan. Data yang diperoleh merupakan data hasil evaluasi kinerja vendor dari tahun 2021 – 2022 yang terdiri dari 770 dan 9 variabel, diantaranya variable SAP, vendor, nilai awal, prestasi, pelanggaran, total nilai, spph menawarkan, dan spph ditawarkan.

Tabel 3. 2 Dataset Evaluasi Vendor

No	SAP	VENDOR	AWAL	PRE STAS I	PELAN GGARA N	TOTAL NILAI	SPPH DITAWARKA N	SPPH MENAWARK AN
1	0010000222	PT.A	75	1	0	78	95	100
2	0010001079	PT.AAS	75	1	0	78	11	454
3	0010000179	PT.ADH	75	0	0	75	77	644
4	0010000208	PT.ADT	75	4	0	80	202	622
5	0010001201	PT.ARS	75	0	0	75	126	574
6	0010000195	PT.AI	75	0	0	75	27	614
7	0010000722	PT.A	75	5	0	80	28	720
8	0010001240	PT.AK	75	0	0	75	47	441
9	0010000951	PT.AT	75	0	0	75	48	88
.
768	0010001019	PT.ZKS	75	0	0	75	4	3
769	0010000017	PT.ZGA	75	0	-3	72	4	3
770	0010000097	PT.YWI	75	0	-3	72	4	3

1.1.2. Implementasi Algoritma

Sebelum dilakukan implementasi algoritma terlebih dahulu dilakukan tahap *Preprocessing* data. *Preprocessing* data terdiri dari data *cleaning*, seleksi data, dan *normalisasi* data.



Gambar 3. 2 Tahapan *Preprocessing*

Berdasarkan gambar 3.2 merupakan tahapan *Preprocessing* data yang terdiri dari data *cleaning*, *seleksi data*, *normalisasi*, dan implementasi algoritma.

1. *Data cleaning*

Pada tahap pembersihan data (*data cleaning*), dilakukan proses untuk menangani nilai yang hilang, *noise*, dan data yang tidak konsisten.

2. *Seleksi data*

Penyeleksian data dilakukan karena data yang didapatkan tidak semuanya digunakan dalam penelitian ini. Penyeleksian data tujuannya untuk menghilangkan variable data yang tidak diperlukan, ada beberapa atribut yang tidak relevan atau tidak dibutuhkan. Pada penelitian ini terdapat tiga variabel yang digunakan diantaranya variable total nilai, spph ditawarkan dan spph menawarkan.

Tabel 3. 3 Seleksi Data

TOTAL NILAI	SPPH DITAWARKAN	SPPH MENAWARKAN
78	95	100
78	11	454
75	77	644
80	202	622

TOTAL NILAI	SPPH DITAWARKAN	SPPH MENAWARKAN
75	126	574
75	27	614
80	28	720
75	47	441
75	48	88
.	.	.
75	4	3
72	4	3
72	4	3

1. *Outliers*

Outliers adalah data yang berada di luar kisaran umum dari sebagian besar data dalam kumpulan data. Mereka muncul sangat jarang dan berbeda dari data lainnya yang diamati. Dalam penelitian ini, teknik *Inter Quartile Range* (IQR) akan digunakan dengan bantuan visualisasi data dalam bentuk *boxplot*. IQR dihitung dengan mengurangkan kuartil pertama (Q1) dari kuartil ketiga (Q3). (Arif, n.d.).

2. *Normalisasi*

Tabel 3.3 menunjukkan nilai asli dari dataset evaluasi vendor sebelum diproses. Dataset tersebut kemudian diubah menggunakan metode *normalisasi min-max* dengan memproses nilai minimum dan maksimum dari setiap atribut. Rentang yang digunakan dalam metode ini 0 hingga 1.

Berikut tahapan perhitungan *min-max normalization* menggunakan persamaan (7).

$$x^1 = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (7)$$

Sebagai contoh perhitungan untuk metode *min-max normalization* di peroleh dengan menggunakan persamaan (7).

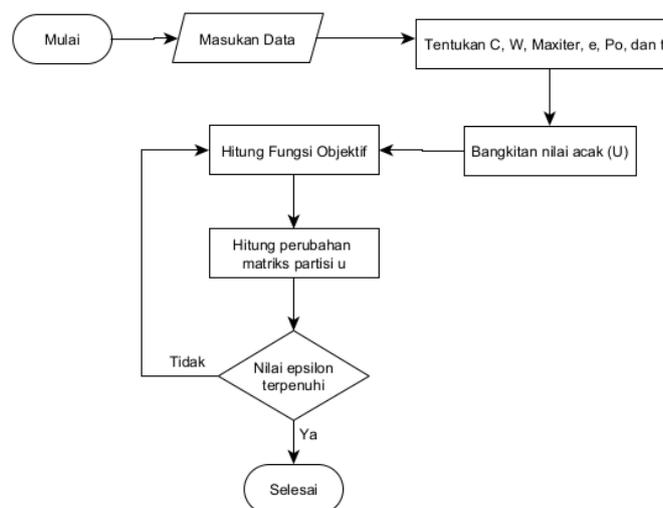
$$95 = \frac{95 - 0}{535 - 0} = \frac{19}{107} = 0,17757$$

Hasil perubahan nilai dengan metode ini dapat dilihat pada Tabel 3.4 Nilai yang dihasilkan setelah pengolahan memiliki rentang nilai yang seimbang.

Tabel 3.4 *Preprocessing*

TOTALNILAI	SPPHDITAWARKAN	SPPHMENAWARKAN
0.061538	0.177570	0.138889
0.061538	0.020561	0.630556
0.053846	0.143925	0.894444
0.066667	0.377570	0.863889
0.053846	0.235514	0.797222
.	.	.
0.058974	0.007477	0.004167
0.053846	0.007477	0.004167
0.053846	0.007477	0.004167
0.046154	0.007477	0.004167
0.046154	0.007477	0.004167

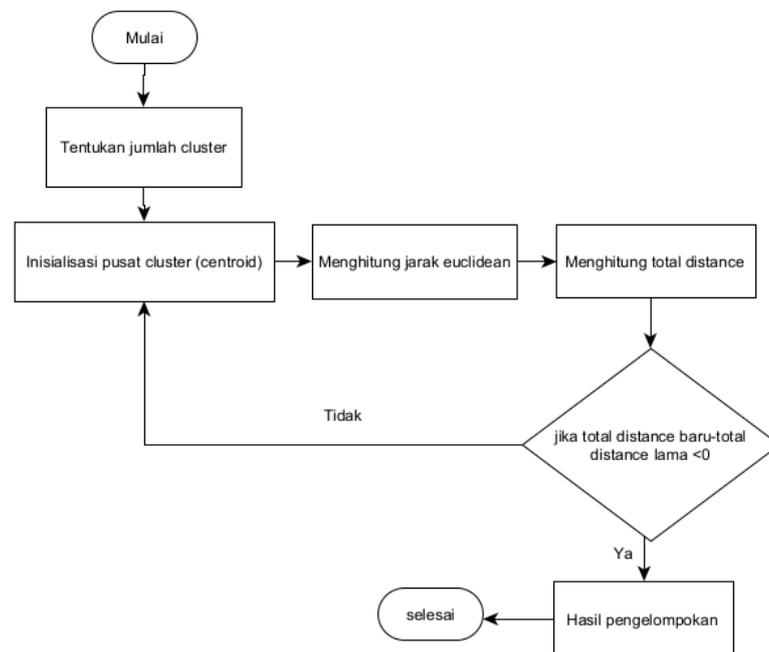
Setelah dilakukan tahap *Preprocessing* selanjutnya melakukan perhitungan menggunakan algoritma *Fuzzy C-Means*. Kemudian di lanjut perhitungan algoritma *K-Medoid* dengan menggunakan bahasa pemrograman *python*. Berikut tahapan dengan menerapkan algoritma *Fuzzy C-means* dapat dilihat pada Gambar 3.3.



Gambar 3.3 Flowcart *Fuzzy C-Means*

Pada metode pengklasteran data menggunakan algoritma *Fuzzy C-Means*, langkah awal yang dilakukan adalah memasukkan data yang akan diklaster dalam bentuk matriks $n \times m$. Kemudian, beberapa indikator yang diperlukan dalam metode *Fuzzy C-Means* harus ditentukan. Setelah itu, bilangan acak digunakan untuk membangkitkan pusat klaster data. Pusat klaster data dihitung, fungsi objektif pada setiap iterasi dihitung, dan perubahan matriks partisi dihitung. Kemudian, kondisi berhenti diperiksa dengan memeriksa apakah nilai epsilon, yang merupakan salah satu indikator, telah terpenuhi atau belum. Jika sudah terpenuhi, iterasi dihentikan. Jika iterasi telah mencapai batas maksimal, kondisi berhenti juga dicapai.

Kemudian untuk tahapan menggunakan algoritma *K-medoids* yaitu dapat dilihat pada Gambar 3.4.



Gambar 3. 4 Flowchat Algoritma K-Medoids

Gambar 3.4 menjelaskan langkah-langkah algoritma *K-medoids* yang terdiri dari beberapa tahap. Tahap pertama adalah menentukan nilai K

yang akan menjadi jumlah *cluster* yang dibentuk. Selanjutnya, setiap data atau objek akan dialokasikan ke dalam *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidean Distance*. Kemudian, dari setiap *cluster* akan dipilih secara acak satu objek sebagai *medoid* awal. Selanjutnya, akan dihitung jarak setiap objek pada masing-masing *cluster* dengan *medoid* baru yang dipilih. Selanjutnya, akan dihitung total simpangan dengan menghitung selisih antara total jarak baru dan total jarak lama. Jika nilai simpangan lebih kecil dari nol, maka objek pada *cluster* tersebut akan ditukar dengan data *cluster* untuk membentuk sebuah *medoid* baru. Tahap ini akan diulangi hingga tidak terjadi perubahan *medoid*. Sehingga, pada akhirnya, akan didapatkan hasil klasterisasi dengan *cluster-cluster* dan anggota-anggota pada masing-masing *cluster*.

1.1.1. Evaluasi

Pada penelitian ini, evaluasi *clustering* akan dilakukan menggunakan *Silhouette Coefficient*. Setelah mendapatkan hasil perhitungan *data mining* yang sesuai, akan dilakukan evaluasi menggunakan *Silhouette Coefficient* untuk menilai validitas algoritma dalam mengelompokkan *cluster*. *Silhouette Coefficient* memiliki rentang nilai antara -1 hingga 1, dimana semakin mendekati 1 menunjukkan bahwa pengelompokan data dalam satu *cluster* semakin baik. Sebaliknya, semakin mendekati -1 menunjukkan bahwa pengelompokan data dalam satu *cluster* semakin buruk..

Silhouette coefficient digunakan untuk mengevaluasi sejauh mana suatu objek ditempatkan dengan baik dalam suatu *cluster*. Langkah-langkah perhitungan yang digunakan dalam *Silhouette coefficient* adalah sebagai berikut (Wira et al., 2019) :

1. Menghitung rata-rata jarak dari suatu dokumen misalkan *i* dengan semua dokumen lain yang berada dalam satu *cluster*, dengan persamaan (8).

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (8)$$

dengan *j* adalah dokumen lain dalam satu *cluster* *A* dan *d(i,j)* adalah jarak antara dokumen *i* dengan *j*.

1. menghitung rata-rata jarak dari dokumen i tersebut dengan semua dokumen di *cluster* lain, dan diambil nilai terkecilnya, menggunakan persamaan (9).

$$d(i, C) = \frac{1}{|A|} \sum_{J \in C} d(i, j) \quad (9)$$

dengan $d(i, C)$ adalah jarak rata-rata dokumen i dengan semua objek pada *cluster* lain C dimana $A \neq C$, dapat dihitung menggunakan persamaan (10).

$$b(i) = \min_{C \neq A} d(i, C) \quad (10)$$

2. Nilai *Silhouette Coefficient* menghitung menggunakan persamaan (11) :

$$s(i) = \frac{d(i) - a(i)}{\max(a(i), b(i))} \quad (11)$$

